Taylor & Francis
Taylor & Francis Group

Coastal Engineering Journal

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tcej20

A metastatistical frequency analysis of extreme storm surge hazard along the US coastline

Georgios Boumis, Hamed R. Moftakhari & Hamid Moradkhani

COASTAL ENGINEERING JOURNAL
2024, VOL. 66, NO. 2, 380–394
https://doi.org/10.1080/21664250.2024.2338323

A metastatistical frequency analysis of extreme storm surge hazard along the US coastline

Georgios Boumis [a,b], Hamed R. Moftakhari [a,b] and Hamid Moradkhani [a,b]

[a]Center for Complex Hydrosystems Research, University of Alabama, Tuscaloosa, AL, USA; [b]Department of Civil, Construction, and Environmental Engineering, University of Alabama, Tuscaloosa, AL, USA

ABSTRACT
Frequency analysis of extreme storm surge is crucial for coastal flood risk assessments. To date, such analyses are based on traditional extreme value theory (EVT) and its associated generalized extreme value (GEV) distribution. The metastatistical extreme value distribution (MEVD) provides a new approach that can alleviate limitations of EVT. This paper provides a comparison between the GEV distribution and the MEVD on their ability to predict "unseen" upper-tail quantiles of storm surge along the US coastline. We analyze the error structure of these distributions by performing a cross-validation experiment where we repeatedly divide the data record into a calibration and validation set, respectively, and then compute the predictive non-dimensional error. We find that the MEVD provides comparable estimates of extreme storm surge to those of the GEV distribution, with discrepancies being subtle and dependent on tide gauge location and calibration set length. Additionally, we show that predictions from the MEVD are more robust with less variability in error. Finally, we illustrate that the employment of the MEVD, as opposed to classical EVT, can lead to remarkable differences in design storm surge height; this has serious implications for engineering applications at sites where the novel MEVD is found more appropriate.

ARTICLE HISTORY
Received 12 June 2023
Accepted 28 March 2024
Published online

KEYWORDS
Coastal hazards; extreme storm surge; extreme value theory; metastatistics

1. Introductory information

Coastal zones around the globe are vulnerable to floods because extreme weather conditions, e.g. (extra)tropical cyclones, can induce intense storm surges and significantly raise the sea water level (Pugh and Woodworth 2014; Wahl et al. 2017). Particularly in the United States (US), tropical cyclone-generated storm surges are among the costliest and deadliest natural hazards (Needham and Keim 2012, 2014). As an example, the total economic losses following the aftermath of Hurricanes Harvey and Irma (2017) exceeded $120 billion, while more than 100 people lost their lives after the two hurricanes made landfall (Klotzbach et al. 2018).

More recently, Hurricane Ian (2022) unleashed catastrophic storm surges leaving coastal communities in southwestern Florida devastated. Coastal flood risk is usually measured as the product of the probability of occurrence of an extreme storm surge event and the subsequent flooding consequence which, in turn, is the product of exposure and vulnerability to flooding (Hawkes et al. 2008). Thus, accurate computation of event occurrence probability through extreme value analysis plays a key role in risk assessments related to coastal hazards. This type of analysis becomes increasingly important considering that more than 600 million people, worldwide, reside currently in low-lying coastal areas with this number only expected to rise in the future (Kulp and Strauss 2019; Neumann et al. 2015). However, estimation of probabilities associated with the upper tail of a statistical distribution is challenging. Extreme storm surges are, by definition, highly unlikely to occur, while the observational record is most of the time considerably shorter than the return period of interest, e.g. from 50 years for the construction of a breakwater to even 10,000 years for the design of a nuclear power plant. As a result, uncertainty in upper-tail quantile estimates can be extremely high (Lin and Emanuel 2016).

Frequency analysis of extreme storm surge has been historically based on the well-grounded extreme value theory (EVT) (S. G. Coles and Tawn 1994; S. Coles 2001; Reiss and Thomas 1997). The latter, represents a sound theoretical framework to calculate the probability of occurrence of a variable at events of extremely high magnitude and has been used in many studies concerning other environmental processes as well, e.g. precipitation (Emmanouil et al. 2020; Hanel and Adri Buishand 2010), river discharge (Villarini and Smith 2010; Villarini et al. 2011) and wind (Alaya, Zwiers, and Zhang 2021; Fawad et al. 2019), among others. EVT identifies the generalized extreme value (GEV) distribution as a suitable candidate distribution

CONTACT Georgios Boumis gboumis@crimson.ua.edu
Supplemental data for this article can be accessed online at https://doi.org/10.1080/21664250.2024.2338323

for fitting maxima of observations (or model simulations) within blocks. Typically, these blocks are assumed to be time periods of 1-year length with the sampling scheme then referred to as annual maxima (AM) (S. Coles 2001). In its original form, the GEV distribution arises as an asymptotic distribution for the maximum of a sequence of n independent and identically distributed (*i.i.d.*) random variables, while it also assumes that $n \to \infty$. Earlier studies have shown that the GEV model can reliably estimate extreme storm surge in many regions around the world, e.g. Argentina (D'Onofrio, Fiore, and Romero 1999), western Europe (Hamdi et al. 2014), and southeastern Asia (Cid et al. 2018). In the US, specifically, FEMA (Federal Emergency Management Agency) recommends the GEV distribution for modeling extreme storm surge along the Pacific Coast (Wallace et al. 2005), while application of the model to extreme storm surge along the Gulf of Mexico and the Northwest Atlantic Ocean suggests that the GEV distribution can reproduce upper-tail quantiles of storm surge (Bernier and Thompson 2006; Boumis, Moftakhari, and Moradkhani 2023b; Huang, Xu, and Nnaji 2008; Rashid, Moftakhari, and Moradkhani 2024). Nevertheless, there exist two fundamental issues with the formulation of the GEV model: 1) it does not maximize the benefit of all available information as it merely uses the AM and discards most observations including other extremes that may occur within a single year (Volpi et al. 2019), and 2) it presumes that the number of independent events per year becomes very large, i.e. $n \to \infty$ (De Haan, and Ferreira 2006), as mentioned earlier. Considering these limitations, both practical and theoretical, it seems tempting to investigate possible alternatives that relax the latter assumption, while simultaneously take advantage of the entire observational record of possibly high frequency.

A different approach that overcomes the limitations of the traditional GEV distribution is the doubly stochastic and non-asymptotic distribution, first introduced by Marani and Ignaccolo (2015), referred to as the metastatistical extreme value distribution (MEVD) (Marani and Ignaccolo 2015; Marani and Zorzetto 2019). This type of distribution treats both the parameters (θ) and the number of events per year (n) as random variables, a concept also known as superstatistics (Beck and Cohen 2003). The metastatistical extreme value (MEV) model is capable of estimating upper-tail quantiles by utilizing non-extreme independent events, i.e. by making use of all the available data which can be assumed unrelated, in contrast to the GEV distribution which relies only on annual maximum values. The MEVD has been previously applied to fluvial flood frequency analysis in the US (Basso et al. 2021; Miniussi, Marani, and Villarini 2020) and Germany (Mushtaq et al. 2022), recurrence of rainfall extremes over Austria (Schellander, Lieb, and Hell 2019) and the US (Marra et al. 2018; Zorzetto and Marani 2020), and hurricane intensities over the Atlantic Ocean (Hosseini, Scaioni, and Marani 2020). More recently, Caruso and Marani (2022) applied the MEVD to the combined effect of storm surge and tide, i.e. the storm-tide in Europe, whereas Vidrio-Sahagún and He (2022) developed an explicit non-stationary formulation of the MEVD for rainfall extremes in Mexico. The current literature indicates that the MEVD provides at least comparable upper-tail quantile estimates and reduces predictive uncertainty when compared to the traditional GEV approach, especially when the recurrence period of interest is substantially longer than the available observational record.

To the best of our knowledge, however, the MEVD has yet to be tested on the frequency analysis of extreme storm surge hazard, let alone in coastal regions of the US. This study aims therefore to examine the metastatistics of extreme storm surge along the US coastline, by comparatively analyzing the overall suitability and predictive performance of both distributions (MEV and GEV) to observed and unobserved events, respectively. To this end, we first obtain still water level (SWL) measurements from NOAA (National Oceanic and Atmospheric Administration) tide gauge stations with a long observational record and conduct harmonic analysis to predict the deterministic tidal component of SWL. We then compute the meteorologically induced stochastic surge component as the non-tidal residual. Subsequently, we analyze the characteristics of independent non-extreme events of positive surge, useful for the MEVD, and then evaluate how well the two contrasting distributions fit all available data of each tide gauge with the use of quantile-to-quantile plots. We also examine extreme storm surge predictive error and uncertainty associated with prediction of the next anticipated extreme event that has not yet been observed, by means of a cross-validation exercise in which we repeatedly divide the available data record into two distinct sets: 1) a short "gauged" calibration set to be used for parameter estimation and 2) a long "ungauged" validation set employed for error estimation. Lastly, we demonstrate that opting for the MEVD over classical EVT might lead to substantial differences in design storm surge height, important for engineering practices, particularly noteworthy in locations where the novel MEVD is deemed more appropriate than the traditional GEV distribution. In summary, this work, relevant for coastal engineering applications, contributes to existing literature as it extends the MEV framework, for the first time, to the meteorological component of SWLs. In contrast to earlier works, it constitutes the first attempt to apply the MEV model to storm surge data of multiple coasts around the US, while it also underscores and showcases the importance of tailoring the MEVD at each tide gauge location individually, in particular with respect to modeling of non-extreme events. In addition, it highlights practical implications of choosing this fairly new approach of extreme value analysis over traditional EVT. Along the way, as opposed to previous similar studies, we also question the use of a generalized Pareto (GP) distribution within the context of cross-validation. This research article is organized as follows: Section 2 introduces the data, pre-processing methods, and statistical tools used for this work, while our findings are presented in Section 3 along with a discussion of the results. Finally, our study is concluded in Section 4.

2. Dataset & statistical methodology

2.1. Storm surge data & temporal trends

We selected 12 tide gauge stations which are distributed along the US coastline and represent a variety of geographical regions, coastal morphologies, and storm surge regimes (Figure 1). These stations measure the hourly SWL, i.e. the combination of mean sea level (MSL), tide, and non-tidal residual (Serafin, Ruggiero, and Stockdon 2017), and have recorded data for >75 years (Table 1). These data were retrieved from *https://tidesandcurrents.noaa.gov* with the use of the *noaaoceans* R package. For our analysis, we discarded years for which $\geq 20\%$ of the data were missing. To obtain

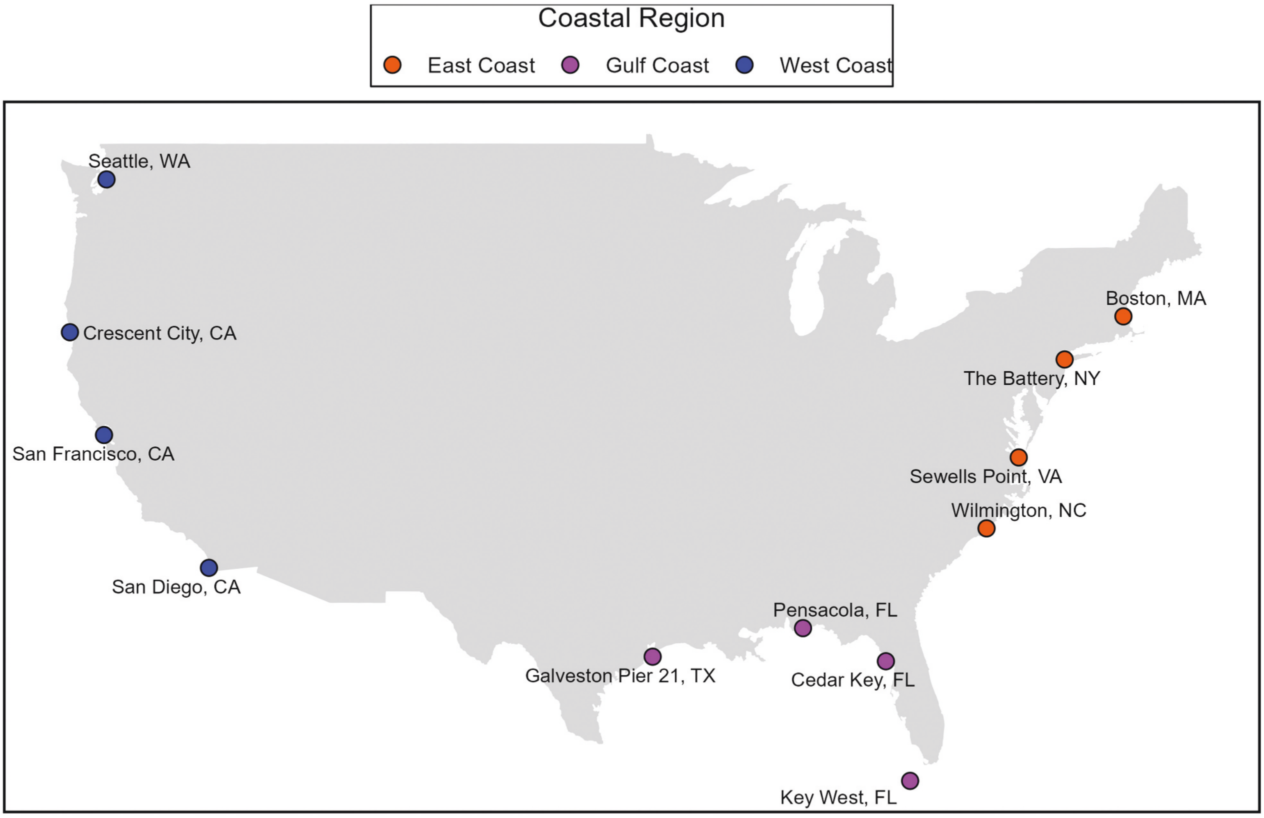

Figure 1. Distribution of selected tide gauges along the US coastline which are analyzed in this study. Orange color indicates stations of East Coast, whereas those of Gulf Coast and West Coast are presented with purple and blue color, respectively.

Table 1. Geographical information, length of record after data pre-processing (years), and average number of independent positive surge events (per year) for each tide gauge station under consideration.

Station Name	State	Longitude	Latitude	Length of Record	# Independent Events
Boston	MA	−71.05	42.35	99	~ 70/yr
The Battery	NY	−74.01	40.70	91	~ 69/yr
Sewells Point	VA	−76.33	36.95	92	~ 64/yr
Wilmington	NC	−77.95	34.22	85	~ 73/yr
Key West	FL	−81.81	24.55	106	~ 65/yr
Cedar Key	FL	−83.03	29.14	94	~ 68/yr
Pensacola	FL	−87.21	30.40	95	~ 62/yr
Galveston Pier 21	TX	−94.79	29.31	107	~ 67/yr
San Diego	CA	−117.17	32.71	112	~ 69/yr
San Francisco	CA	−122.47	37.81	122	~ 68/yr
Crescent City	CA	−124.18	41.75	76	~ 66/yr
Seattle	WA	−122.34	47.60	123	~ 67/yr

the non-tidal residual, i.e. a reasonable proxy for the meteorologically induced surge (Nasr et al. 2021), we performed tidal harmonic analysis on a year-by-year basis using the *TideHarmonics* R package. Specifically, for each year, we fitted 60 major tidal constituents and predicted the tidal water level; the surge was then computed by subtraction of predictions from the observations of SWL. By conducting harmonic analysis on a rolling-year basis we discount the effect of MSL (involving its seasonal cycle) on the estimated surge component, and thus our non-tidal residuals are not affected by potential sea-level rise (Wahl et al. 2015). Finally, for the tide gauge at Crescent City in California, we disregarded the anomalously high (>2.50 m) maximum non-tidal residual in the year 2011 that is associated with the Great East Japan Earthquake and the resultant tsunami, thus not driven by meteorological factors.

We examined any possible trends in storm surge AM by means of the Mann-Kendall non-parametric test, which is based on the correlation between ranks of earlier and later values of a time series (Collaud Coen et al. 2020; Wang et al. 2020). The Mann–Kendall test statistic (S) can be assumed as normally distributed for a sample size of $n \geq 8$ (Mann 1945). Thus, inference on the null hypothesis (H_0) of no trend follows then by comparing the standardized normal statistic ($S_{std} \sim \Phi[0, 1]$) with the theoretical value $\Phi_{1-\alpha/2}$, where α is the significance level (in this case, 5%). The H_0 hypothesis cannot be rejected if $p_{value} > 0.05$.

2.2. Traditional extreme value theory

For a given sample of storm surge AM, which can be assumed independent, classical EVT recognizes the GEV distribution as an appropriate non-degenerate and asymptotic distribution, expressed as:

$$F(z; \mu, \sigma, \xi) = e^{-\left(1+\xi\frac{(z-\mu)}{\sigma}\right)^{-\frac{1}{\xi}}} \tag{1}$$

where F stands for cumulative distribution function, μ is the location parameter, σ denotes the scale parameter, and ξ is the shape parameter. Depending on the numerical value of ξ, the behavior of the distribution's tails changes according to three types: a) the heavy-tailed case (Fréchet type) for $\xi > 0$, b) the light-tailed case (Gumbel type) when $\xi \rightarrow 0$, and c) the short-tailed case (Weibull type) if $\xi < 0$. These different cases are summarized in the 3-parameter distribution of Equation 1. For this study, the GEV distribution parameters were obtained with the method of L-moments (Hosking 1990) by using the *extRemes* R package. When dealing with small sample sizes (see Subsection 2.5), parameter estimation via the L-moments method provides better estimates and is less sensitive to outliers compared to other approaches, e.g. Maximum Likelihood Estimation (MLE) (Martins and Stedinger 2000). Knowledge of the parameters allows then for retrieving quantiles of extreme storm surge associated with a specific probability of exceedance, denoted by p, using the following equations:

$$z_p = \mu - \frac{\sigma}{\xi}\left(1 - (-log(1-p))^{-\xi}\right), \quad \xi \neq 0 \tag{2}$$

or,

$$z_p = \mu - \sigma log(-log(1-p)), \quad \xi \longrightarrow 0 \tag{3}$$

In common terms, z_p is the return level with probability of occurrence p in any given year and is associated with the return period $T = 1/p$. For more details on traditional EVT and the formulation of the GEV distribution, the reader is referred to Coles (2001).

2.3. Metastatistical extreme value distribution

The MEVD can be utilized in order to model the cumulative distribution of storm surge AM by using a greater number of independent events, the so-called "ordinary" events, than just merely the yearly maximum values. It is expressed in terms of the compound probability:

$$G(z) = \sum_{n=1}^{+\infty} \int_{\Omega_\theta} F(z; \theta)^n g(n, \theta) d\theta \tag{4}$$

where $g(n, \theta)$ is the joint probability distribution of the number of events per year (n) and the parameters (θ), while Ω_θ is the parameter population space. An approximate interpretation of the MEVD follows then as the sample average of yearly distributions of ordinary events:

$$G(z) \cong \frac{1}{M} \sum_{j=1}^{M} F(z; \theta_j)^{n_j} \tag{5}$$

where M is the number of years in the observational record, $F(z; \theta_j)$ is the "ordinary" cumulative distribution function of the j^{th} year, and n_j is the number of ordinary independent events in year j. It is to be noted that θ_j might be estimated on a time window which is longer than 1 year. Specifically, when n_j is small, it seems beneficial to estimate the parameters of $F(z)$ based on multiple years (e.g. blocks of 5 years). For practical applications, the desired storm surge quantile can be obtained numerically from inversion of Equation 5.

A necessary assumption of the MEVD is that ordinary events within a year are independent. To ensure independence, we considered surge values that are positive and are separated by a 3-day lag window. A separation threshold of 3 days has shown to be sufficient for achieving independence between two consecutive surge events (Bernardara et al. 2014; Cid et al. 2016). Practically, for each year, we began by selecting the maximum surge value and then discarded all observations which are within ±3 days of the time associated with the selected maximum; we repeated this process until the last positive surge value had been extracted. Here, we restrict ourselves to positive surges only, since we are interested in quantifying extreme storm surge quantiles that can potentially cause coastal flooding. Hence, negative surges, which are more relevant to navigation issues rather than flooding (D'Onofrio, Fiore, and Pousa 2008), are not taken into account. It is evident that this procedure leads to a much greater number of events (including the annual maximum) per year, as opposed to the AM sampling approach.

To identify the most suitable distribution for ordinary events, we compared three different non-extreme value distributions appropriate for modeling skewed data, namely: a) Weibull, b) Gamma, and c) LogNormal. Gamma has shown to be the best distribution for flood frequency analysis with the MEVD (Miniussi, Marani, and Villarini 2020), while Weibull has been utilized in many applications of the MEVD to rainfall data (Marra et al. 2018; Zorzetto, Botter, and Marani 2016). Because the average number of ordinary events per year is sufficient for each tide gauge (see Table 1), we obtained parameters θ_j using moment-matching estimation (Gamma and LogNormal) and MLE (Weibull) (Delignette-Muller and Dutang 2015) for each single year. Oppositely, in earlier applications of the MEVD, parameters of the "ordinary" distribution have been estimated primarily in blocks of 5 years with the use of the probability-weighted moment method (Greenwood et al. 1979; Hosking, Wallis, and Wood 1985).

2.4. Quantile-to-quantile graphs

For each tide gauge, in order to inspect the accuracy of the two models in terms of capturing probabilities of observed events, quantile-to-quantile plots can be constructed after first fitting the MEVD and the GEV distribution to the entire data record of each station. In other words, these graphs help us to assess the so-called "goodness-of-fit" of each model. Given the ordered sample of storm surge AM, i.e. $z_1 \leq z_2 \leq \ldots \leq z_n$, an empirical cumulative distribution function following the correct plotting position of Weibull (Makkonen 2008) can be defined as:

$$\tilde{F}(z) = \frac{i}{n+1} \tag{6}$$

where i is the rank of z_i in the ordered sample, and n is the number of years of record, or else, the number of storm surge AM. The quantile-to-quantile graphs consist then of the pairs:

$$\left\{ (\tilde{F}^{-1}(\frac{i}{n+1}), \hat{F}^{-1}(\frac{i}{n+1})) : i = 1, \ldots, n \right\} \tag{7}$$

where $\hat{F}$ is the modeled cumulative distribution function, either by the GEV distribution or the MEVD.

2.5. Cross-validation of upper-tail quantile prediction

Although quantile-to-quantile graphs can provide a useful tool to assess the overall goodness-of-fit of the two distributions for modeling observed storm surge AM, these graphs are not able to capture the predictive error and associated uncertainty that one encounters when trying to estimate out-of-sample extremes. For this reason, we examined the ability of the GEV distribution and the MEVD to predict unobserved extreme storm surges with long return periods of typical interest, by carrying out a cross-validation experiment involving 1000 repetitions. More precisely, in each repetition, we first randomly selected C (= 10, or 30) years as a calibration set to be used for parameter estimation, instead of the full record, while the rest $K = M - C$ years were used for validation, i.e. predictive error estimation. Here, M denotes the number of years in the full record, as in Equation 5. By performing the cross-validation experiment for two different values of C, we can more comprehensively assess the predictive performance of the two models investigating both a very short and a moderate length of record of 10 and 30 years, respectively. We chose to keep the calibration sample size relatively small considering that most tide gauges around the globe have a short observational archive. For example, $\sim$ 80% of the tide gauges which are part of the GESLA (Global Extreme Sea Level Analysis) database Version 3.0 (Haigh et al. 2023) report data for <30 years. Next, we sorted the AM of the validation set in ascending order and assigned an empirical probability of non-exceedance ($\tilde{F}_i$) to each annual maximum (z_i) given by $\tilde{F}_i = i/(K+1)$, where i is the rank of z_i and K is the number of years inside the validation set, similar to Equation 6. Subsequently, we estimated with both the GEV distribution and MEVD those quantiles associated with the maximum return period of the validation set, i.e. $T_{max} = 1/(1 - F_{i_{max}})$. Finally, for both distributions, we calculated the non-dimensional error (*NDE*), also known as relative error; iterative calculation of this metric, through our cross-validation exercise, provided a way of quantifying predictive error variability. The *NDE* was computed using the following expression:

$$NDE^j = \frac{z^j_{est} - z^j_{obs}}{z^j_{obs}} \tag{8}$$

where $NDE \in (-\infty, +\infty)$, while z^j_{obs} and z^j_{est} denote the observed and estimated storm surge of iteration j, respectively. We opted for the *NDE* which has a perfect value of zero (when $z_{est} = z_{obs}$), since it provides a standardized way of assessing predictive accuracy, in contrast to a non-relative error metric, and thus allows for an enhanced communication of our results across various coastal regions that exhibit extremes of different magnitude. As an example, a predictive non-relative error of 50 cm would have a totally different interpretation if it was to be obtained for tide gauges along the West Coast, where storm surges tend to be milder, as opposed to stations along the Gulf Coast which are historically struck by catastrophic tropical cyclones. The *NDE*, on the contrary, facilitates a meaningful intercomparison of our results between various US coasts. In addition, this metric denotes an overestimation or underestimation based on its corresponding sign, i.e. positive or negative, respectively.

Because our cross-validation exercise embraces small-size calibration sets, and due to its inherent stochastic and repetitive nature, we decided to not consider the GP distribution (S. Coles 2001) as part of our comparative analysis, as was done in earlier similar works. The GP distribution relies on a peaks-over-threshold approach (Davison and Smith 1990) where the chosen threshold (u) should be high enough for the EVT assumptions to hold, while simultaneously small enough to provide a sufficient sample size for a robust statistical analysis. However, the selection of such a threshold is a considerably hard task for short datasets (Fukutome, Liniger, and Süveges 2015; Langousis et al. 2016; Solari et al. 2017). Also, since the calibration set is randomly picked in each iteration, application of the GP distribution would imply optimizing u for each repetition, possibly via multiple threshold-detection methods (Langousis et al. 2016), thus rendering the cross-validation exercise a particularly difficult task. We illustrate how a single threshold, u, can be problematic within the context of a cross-validation exercise, by examining mean residual life plots (Davison and Smith 1990). Figure S1 shows the mean residual life plot for Boston, MA, as obtained considering all 99 years of record at site. According to the properties of the GP distribution (Text S1), an optimal u is found to be around 0.63 m. However, for $C = 10$ random years of calibration data (i.e. a random cross-validation iteration), the same u appears to be a bad choice for fitting a GP distribution with; a better threshold is obtained for a value of 0.50 m (Figure S2). This suggests that the use of a single threshold u would result in biased GP quantile estimates, thus unfair comparison with the GEV distribution and the MEVD, while optimization of u for each repetition of the cross-validation exercise would be tedious and practically infeasible.

3. Results & discussion

3.1. Evaluation of temporal trends

For each tide gauge, we extracted the time series of storm surge yearly maxima and examined potential temporal trends. Because storm surge is associated with low-pressure weather systems, especially (extra) tropical cyclones, temporal trends in storm surge AM are indicative of storm activity and can reveal patterns of increasing/decreasing storm intensity over the years. Figures S3 to S5 illustrate the maximum storm surge on an annual basis for East Coast, Gulf Coast, and West Coast, respectively. The tide gauge at Wilmington, NC, appears to exhibit an increasing trend in storm surge AM (Figure S3), whereas a decreasing trend can be observed for the tide gauge at San Diego, CA (Figure S5). No apparent trends can be inferred for the rest of the tide gauges through visual inspection only. Indeed, as part of the Mann-Kendall non-parametric test (Subsection 2.1) we obtained a statistically significant positive and negative S statistic for Wilmington and San Diego, respectively, with $p_{values} < 0.05$. However, the test also detected a statistically significant positive trend for the station at Pensacola, FL ($p_{value} = 0.02$). No trends were detected for the remaining stations ($pvalues > 0:05$). Our finding of statistically insignificant storm surge trends for the majority of stations examined here, is in agreement with the latest literature on the matter for the coastal regions under study (Tadesse et al. 2022). Also, these results support the assertion that trends in extreme SWLs along the US coastline are primarily due to MSL rise and not because of changes in storm activity (Boumis, Moftakhari, and Moradkhani 2023a; Menéndez and Woodworth 2010; Woodworth and Blackman 2004). Statistically significant temporal trends, obtained here for only a handful of tide gauges, call into question the suitability of stationary distributions for modeling extreme storm surge hazard at these specific locations. However, this finding does not have a noteworthy impact on our comparison between the GEV distribution and the MEVD, at least with regard to predictive error estimation, since our cross-validation approach works by randomly selecting only a small number of years for parameter estimation, i.e. calibration set (Subsection 2.5), and thus eradicates any existing trends.

3.2. Assessment of overall goodness-of-fit

Model checking via quantile-to-quantile plots cannot justify extrapolation to "unseen" extremes beyond the length of the observational record. However, it provides a reasonable way of assessing the total goodness-of-fit of a distribution to all available data used to estimate the distribution's parameters (S. Coles 2001). Figures 2 – 4 display quantile-to-quantile graphs for

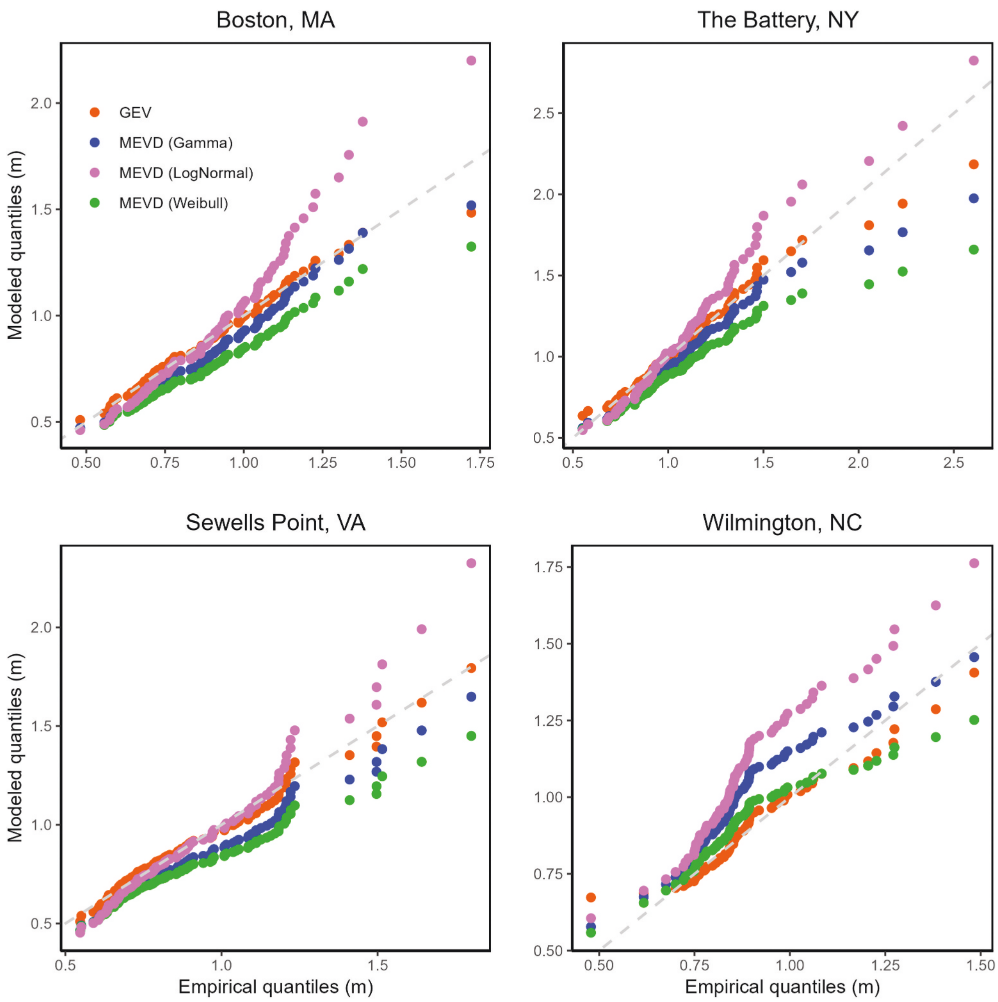

Figure 2. Quantile-to-quantile plots of storm surge annual maxima for tide gauges along the East Coast. These plots are obtained by fitting both distributions to all available data in the observational record of each tide gauge. The GEV model fit is denoted by orange dots, while the MEV model fit is shown with blue (Gamma), pink (LogNormal), and green (Weibull) dots, respectively. The gray dashed line represents the 1:1 line.

stations along the East Coast, Gulf Coast, and West Coast, respectively. These graphs were derived by fitting the GEV distribution using the L-moments method, while for the MEVD, all three possible variants of the "ordinary" distribution were tested here, by employing moment-matching estimation (Gamma and LogNormal) and MLE (Weibull). From Figure 2 it can be inferred that, in general, the two distributions provide an adequate fit to storm surge AM of the studied stations at East Coast. However, it is also apparent that the goodness-of-fit of the MEVD heavily relies on the choice of the distribution used to model ordinary events, a choice which varies with tide gauge location. As an example, Gamma yields the most satisfactory fit of the MEVD for the station at Boston, MA, and The Battery, NY, whereas LogNormal and Weibull are preferred for the tide gauge at Sewells Point, VA, and Wilmington, NC, respectively. From all four panels (Figure 2), it is also evident that both the GEV distribution and the best variant of the MEVD underestimate upper-tail storm surges with the latter distribution yielding higher discrepancies than the former. The underestimations of upper quantiles are more pronounced for the station at The Battery, NY, where observed storm surge AM tend to be greater than that of other stations (see *x* axes in Figure 2). When examining tide gauges along the Gulf Coast (Figure 3), it is shown that at Cedar Key, FL, both the LogNormal variant of the MEVD and the GEV distribution provide a suitable model for the bulk of the empirical data including the right tail. Interestingly, when contrasting LogNormal with Gamma as a suitable model for ordinary events of the MEVD at other tide gauges, i.e. Key West, FL, Pensacola, FL, and Galveston Pier 21, TX, it is clear that the former yields a better fit for upper quantiles than the latter, while it is inferior for lower storm surges. This finding may suggest that the storm surge data along the Gulf Coast are likely to originate from a fusion of storm surge populations (e.g. tropical cyclone-related or not) and might be better modeled by a mixture of distributions. Besides, at the same tide gauges, both the GEV distribution and the MEVD are unable to capture the most extreme observations (Figure 3). Similar to East Coast, the

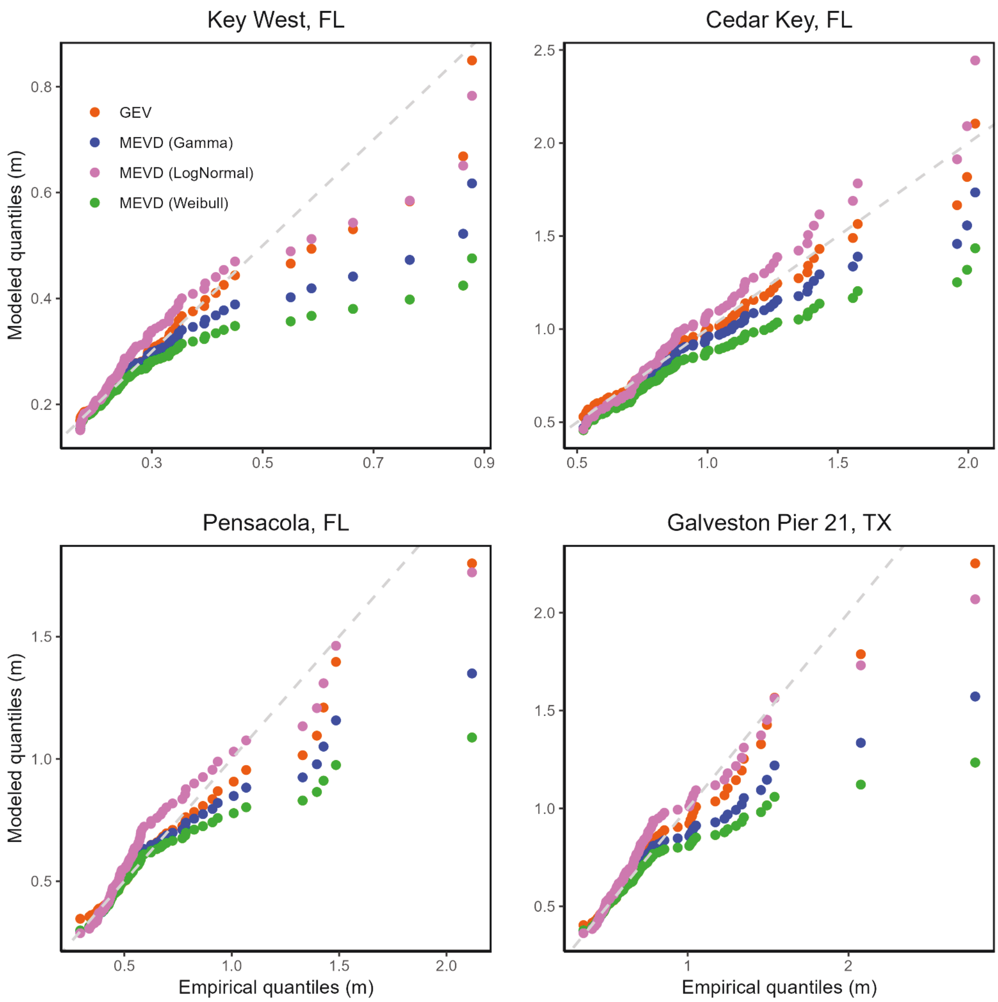

Figure 3. Quantile-to-quantile plots of storm surge annual maxima for tide gauges along the Gulf Coast. These plots are obtained by fitting both distributions to all available data in the observational record of each tide gauge. The GEV model fit is denoted by orange dots, while the MEV model fit is shown with blue (Gamma), pink (LogNormal), and green (Weibull) dots, respectively. The gray dashed line represents the 1:1 line.

goodness-of-fit of the MEVD for stations at West Coast greatly depends upon the choice of the so-called "ordinary" distribution (Figure 4). Even though Gamma appears to work well for the tide gauges at Seattle, WA, and San Francisco, CA, the MEVD fits the observed data better when LogNormal and Weibull are used for San Diego, CA, and Crescent City, CA, respectively. For every tide gauge examined along the West Coast, the best variant of the MEVD displays a satisfactory fit, analogous to that of the GEV distribution, while both models capture most observations quite well without any prominent discrepancies.

3.3. Analysis of predictive error structure

We now assess the ability of the MEVD and the GEV distribution to predict previously unobserved upper-tail quantiles, by exploring the predictive error structure. For all tide gauges under study, we carried out our cross-validation exercise testing different candidate distributions to model ordinary surge events for the MEVD. It is to be noted that the optimal choice of "ordinary" distribution within the cross-validation context was highly station-dependent and therefore a universal choice could not be made; this is analogous to our finding about the overall goodness-of-fit of the MEVD, where all available data were assumed to have been observed and thus used for model fitting (see Subsection 3.2).

The distributions of the *NDE* for tide gauges along the East Coast, Gulf Coast and West Coast are shown in Figures 5–7, respectively; results are shown for both calibration sample sizes (C). These distributions are then summarized by their median values, which are shown in Tables 2 and 3 for $C = 10$ and 30 years, respectively. These tables reveal subtle differences between median absolute errors of the GEV distribution and the best variant of the MEVD for most stations. As an example, for the tide gauge at Boston, MA, both distributions yield a median absolute *NDE* of 0.18 when $C = 10$ (Table 2), while a marginal difference of 0.02 between median absolute *NDEs* is observed in favor of the MEVD when $C = 30$ (Table 3). Similarly, at Cedar Key, FL, the GEV distribution exhibits a slightly better

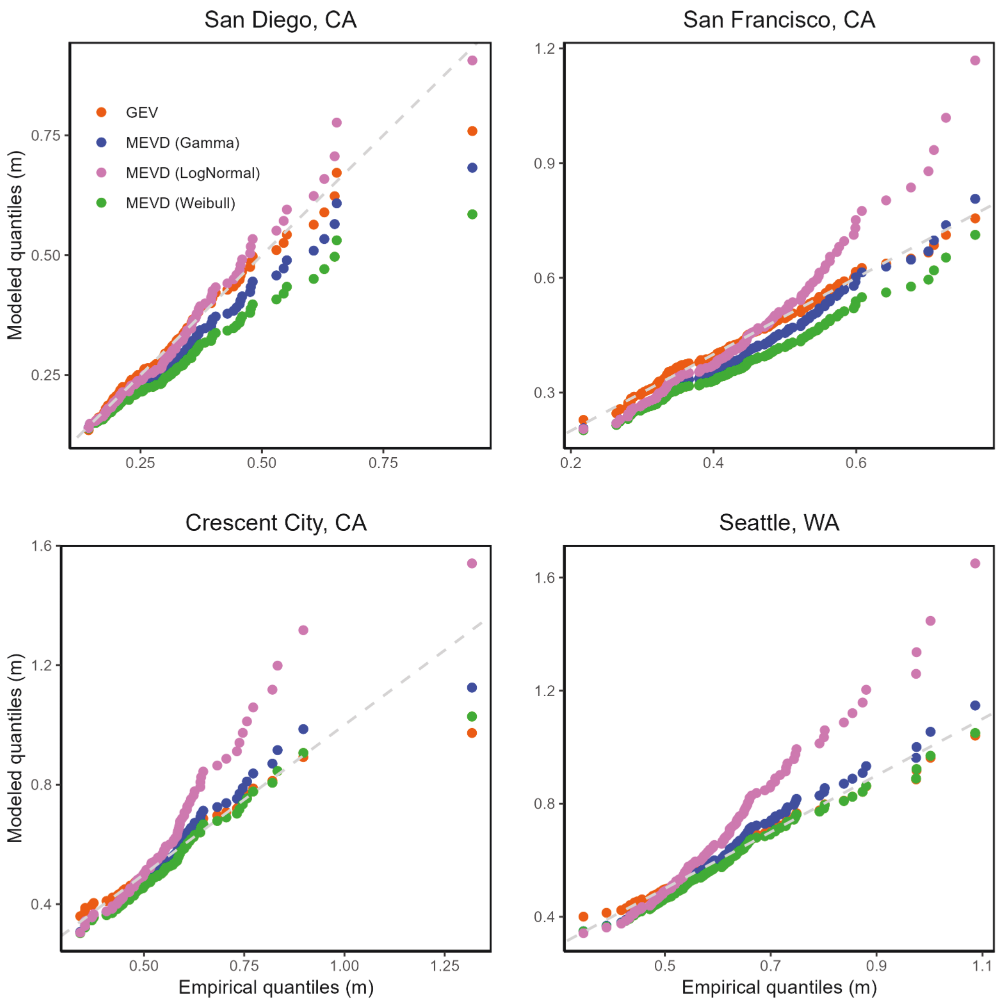

Figure 4. Quantile-to-quantile plots of storm surge annual maxima for tide gauges along the West Coast. These plots are obtained by fitting both distributions to all available data in the observational record of each tide gauge. The GEV model fit is denoted by orange dots, while the MEV model fit is shown with blue (Gamma), pink (LogNormal), and green (Weibull) dots, respectively. The gray dashed line represents the 1:1 line.

median absolute *NDE* of 0.05 than that of the MEVD which is 0.07 when $T_{max} = 84$ (Table 2), whereas both distributions display the same median absolute *NDE* when $T_{max} = 64$ (Table 3). In another instance, at San Francisco, CA, these discrepancies remain quite small, but in favor of the GEV distribution, with values being 0.03 and 0.04 when $C = 10$ (Table 2) and $C = 30$ (Table 3), respectively. It is essential to note that the cross-validation experiment is a stochastic procedure, and thus minor variations in these results can be anticipated when using a different random seed for sampling. Also, different parameter estimation methods may result in dissimilar variances of the parameters, hence using different estimation methods for different distributions may add a source of small, yet potential, discrepancies in median *NDE* estimates and its variance. Therefore, these findings, i.e. the subtle differences observed in the *NDE* distributions of many tide gauges, prevent a definitive assertion regarding which model is more suitable for estimating low-probability unobserved storm surges. Nevertheless, there are some cases where the disparities between the GEV distribution and the optimal version of the MEVD are more noticeable. For instance, at The Battery, NY, the MEV model employing a LogNormal distribution demonstrates superior performance compared to the GEV model, whether with a very brief (see Table 2) or an extended (see Table 3) calibration set. Conversely, at Galveston Pier 21, TX, the GEV distribution produces a significantly smaller nondimensional error, suggesting its superiority over the MEVD.

Besides, it is evident from the cross-validation results that three other remarks can be transparently made. Initially, both distributions consistently result in underestimations when attempting to predict extreme unobserved storm surge events. This pattern holds true for all three coastal regions investigated, regardless of the calibration sample sizes, as evident from the prevalent negative values in Tables 2 and 3. Secondly, the MEVD displays reduced variability in predictive error, evident in narrower boxplots with fewer outliers, compared to the GEV distribution for most tide gauges (see Figures 5–7). The GEV distribution, on the other hand, shows high sensitivity to the subset of data used

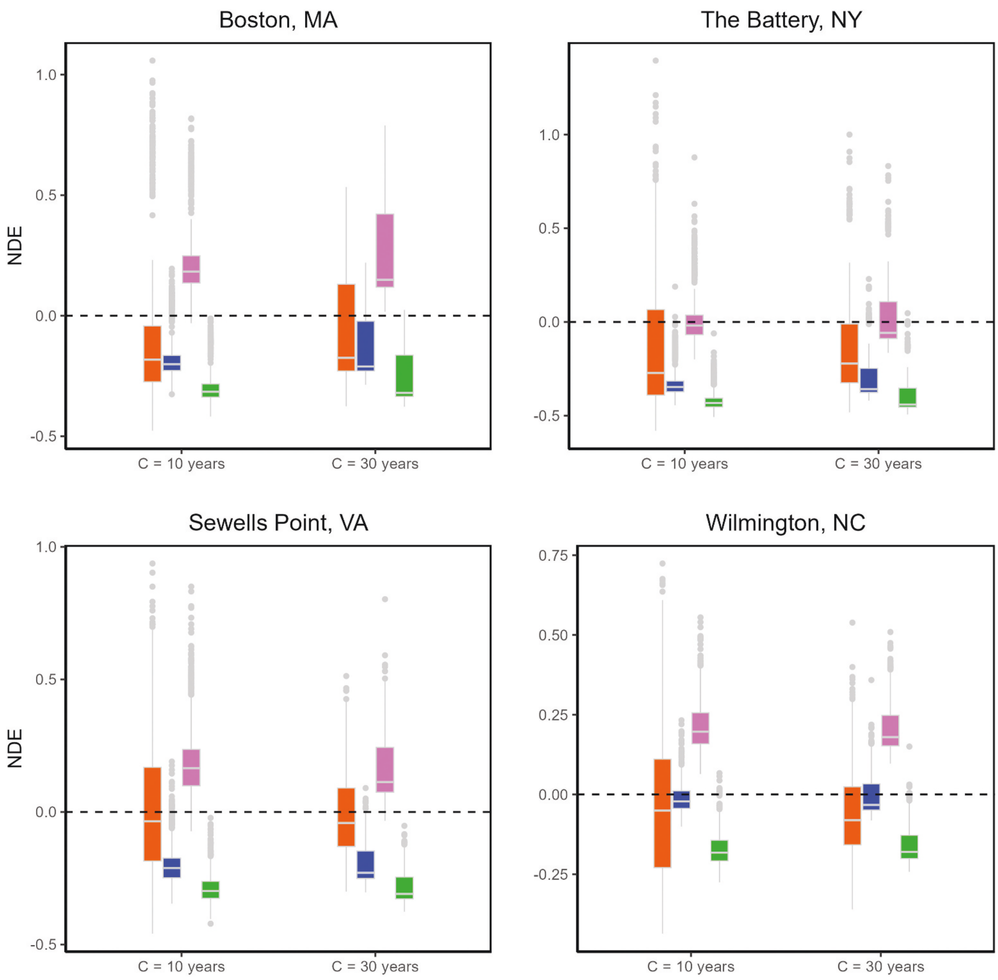

Figure 5. Distribution of the non-dimensional error (NDE) for tide gauges along the East Coast. Results for both calibration sample sizes (*C*), i.e. 10 and 30 years, are shown. Orange boxplots refer to the GEV distribution, while those with blue, pink, and green color denote the MEVD with a Gamma, LogNormal, and Weibull "ordinary" distribution, respectively. The black horizontal dashed line indicates a perfect NDE of zero value.

for model calibration, particularly pronounced when the calibration set length is short, i.e. 10 years, as a consequence of predicting an event with a much longer return period. In conclusion, as highlighted earlier, there is no unanimous agreement on the "ordinary" distribution for the MEV model that could be universally applied across tide gauges in all three coastal regions of the US. Our findings indicate that this decision should be best made on a site-by-site basis to attain optimal performance. For example, at Wilmington, NC, and Seattle, WA, the Gamma distribution demonstrates significantly better predictive performance. Meanwhile, at Key West, FL, and The Battery, NY, the Weibull and LogNormal distribution, respectively, appear to be the most suitable choice for ordinary surge events (Tables 2 and 3). To further conclude, our study underscores the importance of tailoring the choice of the "ordinary" distribution within the MEV framework to the specific characteristics of each tide gauge location. This nuanced understanding, exemplified by the varied performance of different distributions at specific stations, highlights the complexity of modeling storm surge events and emphasizes the need for a location-specific strategy when employing the MEV framework.

3.4. *Implications for design storm surge height*

In the field of coastal engineering, conventional applications of traditional EVT are pivotal. EVT has established an essential framework for the design and evaluation of structures, allowing engineers to take into account uncommon yet potentially catastrophic events unique to coastal settings. This ensures that structures are reliable and resilient enough to endure these rare but impactful incidents. Notably, classical EVT has been prevalent in scientific literature addressing coastal engineering, underlining its established significance in the field (Hawkes et al. 2008). The MEVD, however, stemming from recent advancements in extreme value statistics, presents an alternative approach that may prove more reliable and robust for estimating extreme sea levels. Importantly, this effectiveness can be site-specific, as shown earlier,

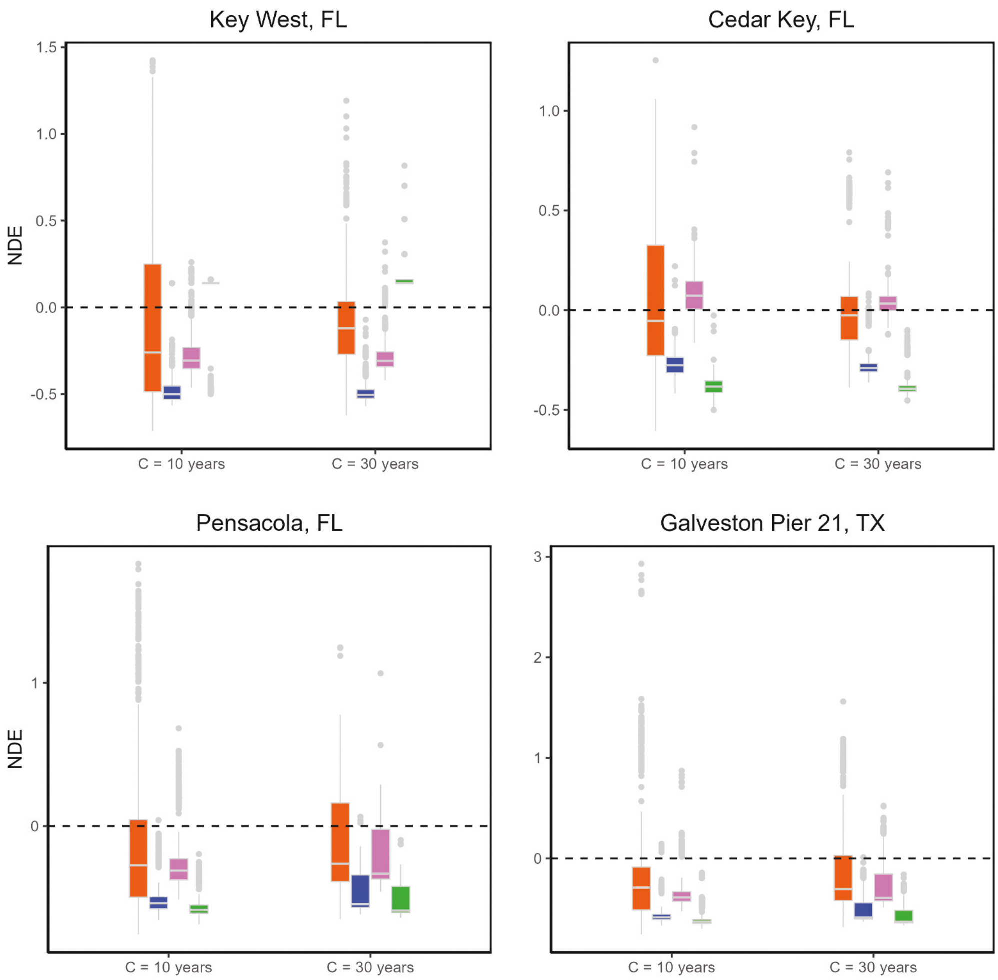

Figure 6. Distribution of the non-dimensional error (NDE) for tide gauges along the Gulf Coast. Results for both calibration sample sizes (*C*), i.e. 10 and 30 years, are shown. Orange boxplots refer to the GEV distribution, while those with blue, pink, and green color denote the MEVD with a Gamma, LogNormal, and Weibull "ordinary" distribution, respectively. The black horizontal dashed line indicates a perfect NDE of zero value.

meaning that the MEV model might be particularly advantageous in certain locations. As an illustration of the consequences of employing the MEVD over the GEV distribution in coastal engineering practices, we focus on prediction of the 100-year storm surge event at The Battery, NY. The findings in Subsection 3.3 revealed that for the tide gauge at The Battery, NY, utilizing the MEVD with a LogNormal distribution results in notably lower errors when extrapolating to out-of-sample extreme events (Tables 2 and 3), i.e. its predictive ability is superior to that of the GEV distribution. Hence, we proceeded to fit both the GEV distribution and the MEVD using the entire 91-year record at the site (Table 1), aiming to predict the storm surge height with an average recurrence period of 100 years using both distributions, and then make a comparison. The GEV distribution resulted in a height of 2.21 m, while the MEVD yielded a height of 2.88 m for the same 100-year return period. This disparity of approximately 65 cm is crucial as it can lead to underestimating the required design specifications, potentially compromising the integrity and performance of coastal defenses. It is also worth mentioning that the storm surge height obtained with the MEVD (2.88 m) falls outside the 95% confidence interval computed with the GEV distribution, which is found to be (1.75 m, 2.84 m). This further underlines the distinct difference in design storm surge height that the two distributions yield. Therefore, choosing between the GEV distribution and the MEVD becomes an important decision when informing construction of reliable coastal structures.

4. Conclusion

In this work, we applied the MEV model to storm surge data along the US coastline. Specifically, we analyzed the goodness-of-fit and predictive performance of the MEVD comparing it with that of classical EVT and its associated GEV distribution. Upon visually inspecting quantile-to-quantile plots, created by fitting both models to the entire data record of each tide gauge, it was observed that both the GEV distribution and the MEVD offer reasonable estimates of observed extreme storm surge events. However, the former generally exhibits smaller underestimations. In our cross-

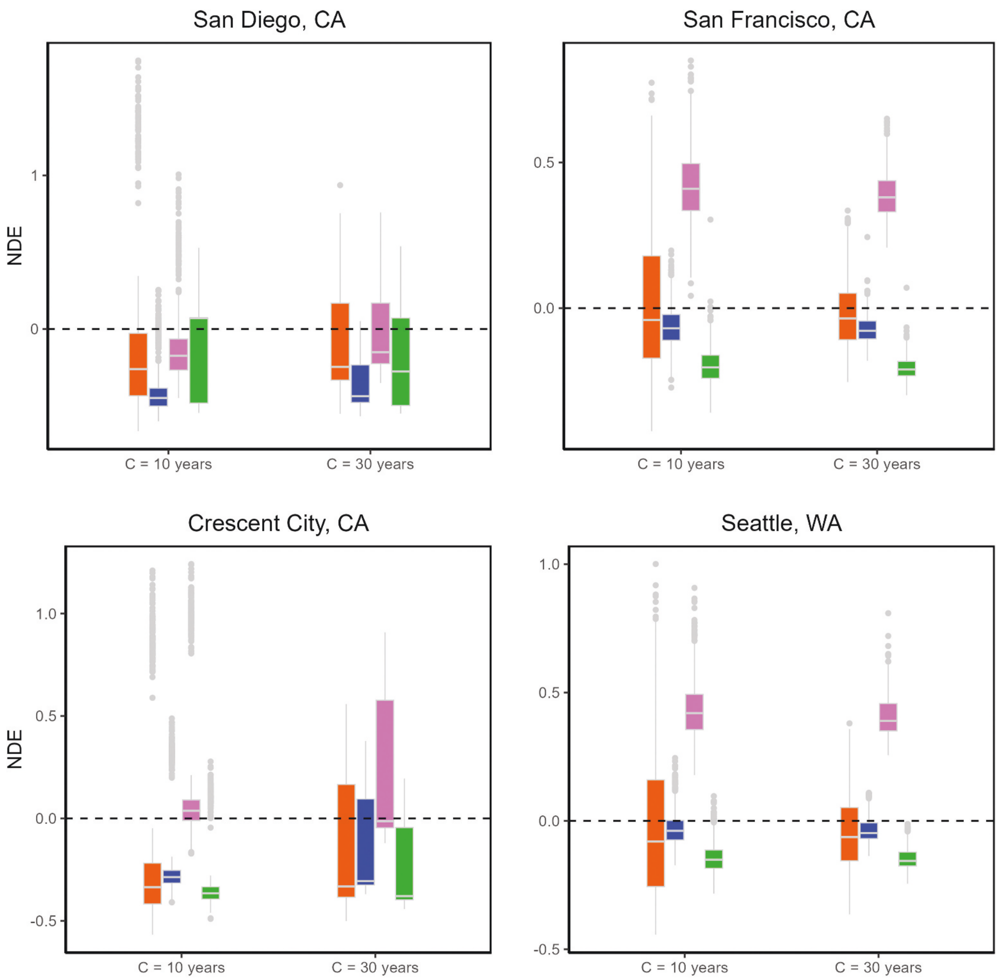

Figure 7. Distribution of the non-dimensional error (NDE) for tide gauges along the West Coast. Results for both calibration sample sizes (*C*), i.e. 10 and 30 years, are shown. Orange boxplots refer to the GEV distribution, while those with blue, pink, and green color denote the MEVD with a Gamma, LogNormal, and Weibull “ordinary” distribution, respectively. The black horizontal dashed line indicates a perfect NDE of zero value.

Table 2. Median value of the nondimensional error (*NDE*) distribution for all tide gauge stations. Results pertain to a calibration sample size of $C = 10$ years, or else, a maximum return period for validation of $T_{max} = M - C$, where M is the data record length of each station. An asterisk indicates the lowest absolute *NDE* for the respective station.

Station Name	GEV	MEVD (Gamma)	MEVD (LogNormal)	MEVD (Weibull)
Boston	−0.18*	−0.20	0.18*	−0.32
The Battery	−0.27	−0.35	−0.02*	−0.43
Sewells Point	−0.04*	−0.21	0.17	−0.30
Wilmington	−0.05	−0.02*	0.20	−0.18
Key West	−0.26	−0.50	−0.31	0.14*
Cedar Key	−0.05*	−0.28	0.07	−0.38
Pensacola	−0.27*	−0.54	−0.31	−0.59
Galveston Pier 21	−0.29*	−0.57	−0.39	0.63
San Diego	−0.26	−0.45	−0.17	0.07*
San Francisco	−0.04*	−0.07	0.41	−0.20
Crescent City	−0.34	−0.29*	−0.34	−0.37
Seattle	−0.08	−0.04*	0.42	−0.15

validation exercise, repeatedly dividing the observational record into a small calibration set and a larger validation set, we observed that the GEV predictions for “unseen” extreme storm surges were unstable. This instability was marked by significant variability in predictive error. However, the MEVD, leveraging enhanced information from ordinary surge events, demonstrated the ability to minimize predictive uncertainty. Specifically, the predicted upper-tail quantiles were found to be robust and less sensitive to the random samples used for calibration. These findings are in-line with earlier comparative studies involving the MEVD (Zorzetto, Botter, and Marani 2016). Overall, our findings indicate that the disparities in the average predictive performance of the two models are often subtle and contingent on factors such as the length of

Table 3. Median value of the non-dimensional error (*NDE*) distribution for all tide gauge stations. Results pertain to a calibration sample size of $C = 30$ years, or else, a maximum return period for validation of $T_{max} = M - C$, where M is the data record length of each station. An asterisk indicates the lowest absolute *NDE* for the respective station.

Station Name	GEV	MEVD (Gamma)	MEVD (LogNormal)	MEVD (Weibull)
Boston	−0.17	−0.21	0.15*	−0.32
The Battery	−0.22	−0.36	−0.06*	−0.44
Sewells Point	−0.04*	−0.23	0.11	−0.31
Wilmington	−0.08	−0.03*	0.18	−0.18
Key West	−0.12*	−0.50	−0.31	0.14
Cedar Key	−0.03*	−0.29	0.03*	−0.39
Pensacola	−0.26*	−0.55	−0.33	−0.59
Galveston Pier 21	−0.31*	−0.59	−0.39	−0.63
San Diego	−0.25	−0.44	−0.15*	−0.28
San Francisco	−0.04*	−0.08	0.38	−0.21
Crescent City	−0.33	−0.31	−0.01*	−0.38
Seattle	−0.06	−0.05*	0.39	−0.16

the calibration set and the location of the tide gauge (Tables 2 and 3). Additionally, our analysis underscores the importance of choosing the "ordinary" distribution for the MEVD on a site-by-site basis for optimal results. In conclusion, we also provided an example illustrating how the newly emerged MEVD can result in significantly different design storm surge heights compared to the well-established GEV distribution. In coastal areas where the MEVD proves to be more effective in predicting out-of-sample extreme events, it is worth re-thinking the use of traditional extreme-value statistical tools.

Acknowledgments

Data sources and R packages used for this analysis have been recognized with in-text references. All authors would like to thank the three anonymous reviewers whose comments helped significantly improve the quality of this work. The first author would also like to acknowledge Ph.D. researcher Da-young Kang for aiding with esthetics of the figures.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This study is funded by the National Science Foundation awards [#2223893 and #2238000].

ORCID

Georgios Boumis http://orcid.org/0000-0001-7825-5239

References

Alaya, MA Ben, F. W. Zwiers, and X. Zhang. 2021. "On Estimating Long Period Wind Speed Return Levels from Annual Maxima." *Weather and Climate Extremes* 34:100388. https://doi.org/10.1016/j.wace.2021.100388.

Basso, Stefano, Gianluca Botter, Ralf Merz, and Arianna Miniussi. 2021. "PHEV! The PHysically-Based Extreme Value Distribution of River Flows." *Environmental Research Letters* 16 (12): 124065. https://doi.org/10.1088/1748-9326/ac3d59

Beck, Christian, and Ezechiel GD Cohen. 2003. "Superstatistics." *Physica A: Statistical Mechanics and Its Applications* 322:267–275. https://doi.org/10.1016/S0378-4371(03)00019-0

Bernardara, P, F. Mazas, X. Kergadallan, and L. Hamm. 2014. "A Two-Step Framework for Over-Threshold Modelling of Environmental Extremes." *Natural Hazards and Earth System Sciences* 14 (3): 635–647. https://doi.org/10.5194/nhess-14-635-2014

Bernier, N. B, and K. R. Thompson. 2006. "Predicting the Frequency of Storm Surges and Extreme Sea Levels in the Northwest Atlantic." *Journal of Geophysical Research Oceans* 111 (C10): 111 (C10. https://doi.org/10.1029/2005JC003168

Boumis, Georgios, Hamed R Moftakhari, and Hamid Moradkhani. 2023a. "Coevolution of Extreme Sea Levels and Sea-Level Rise Under Global Warming." *Earth's Future* 11 (7): e2023EF003649. https://doi.org/10.1029/2023EF003649

Boumis, Georgios, Hamed R Moftakhari, and Hamid Moradkhani. 2023b. "Storm Surge Hazard Estimation Along the US Gulf Coast: A Bayesian Hierarchical Approach." *Coastal Engineering* 185:104371. https://doi.org/10.1016/j.coastaleng.2023.104371

Caruso, Maria Francesca, and Marco Marani. 2022. "Extreme-Coastal-Water-Level Estimation and Projection: A Comparison of Statistical Methods." *Natural Hazards and Earth System Sciences* 22 (3): 1109–1128. https://doi.org/10.5194/nhess-22-1109-2022

Cid, Alba, Melisa Menéndez, Sonia Castanedo, Ana J Abascal, Fernando J Méndez, and Raúl Medina. 2016. "Long-Term Changes in the Frequency, Intensity and Duration of Extreme Storm Surge Events in Southern Europe." *Climate Dynamics* 46 (5): 1503–1516. https://doi.org/10.1007/s00382-015-2659-1

Cid, Alba, Thomas Wahl, Don P Chambers, and Sanne Muis. 2018. "Storm Surge Reconstruction and Return Water Level Estimation in Southeast Asia for the 20th Century." *Journal of Geophysical Research Oceans* 123 (1): 437–451. https://doi.org/10.1002/2017JC013143

Coles, Stuart. 2001. *An Introduction to Statistical Modeling of Extreme Values*. Vol. 208. London: Springer.

Coles, Stuart G, and Jonathan A Tawn. 1994. "Statistical Methods for Multivariate Extremes: An Application to

Structural Design." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 43 (1): 1–31. https://doi.org/10.2307/2986112

Collaud Coen, M, Elisabeth Andrews Martine, Alessandro Bigi, Giovanni Martucci, Gonzague Romanens, Frédéric Vogt, and Laurent Vuilleumier. 2020. "Effects of the Prewhitening Method, the Time Granularity, and the Time Segmentation on the Mann–Kendall Trend Detection and the Associated Sen's Slope." *Atmospheric measurement techniques* 13 (12): 6945–6964. https://doi.org/10.5194/amt-13-6945-2020

Davison, Anthony C, and Richard L Smith. 1990. "Models for Exceedances Over High Thresholds." *Journal of the Royal Statistical Society: Series B (Methodological)* 52 (3): 393–425. https://doi.org/10.1111/j.2517-6161.1990.tb01796.x

De Haan, Laurens, and Ana Ferreira. 2006. *Extreme Value Theory: An Introduction*. Vol. 21. New York: Springer.

Delignette-Muller, Marie Laure, and Christophe Dutang. 2015. "Fitdistrplus: An R Package for Fitting Distributions." *Journal of Statistical Software* 64 (4): 1–34. https://doi.org/10.18637/jss.v064.i04

D'Onofrio, Enrique E, Mónica ME Fiore, and Jorge L Pousa. 2008. "Changes in the Regime of Storm Surges at Buenos Aires, Argentina." *Journal of Coastal Research* 24 (10024): 260–265. https://doi.org/10.2112/05-0588.1

D'Onofrio, Enrique E, Monica ME Fiore, and Silvia I Romero. 1999. "Return Periods of Extreme Water Levels Estimated for Some Vulnerable Areas of Buenos Aires." *Continental Shelf Research* 19 (13): 1681–1693. https://doi.org/10.1016/S0278-4343(98)00115-0

Emmanouil, Stergios, Andreas Langousis, Efthymios I Nikolopoulos, and Emmanouil N Anagnostou. 2020. "Quantitative Assessment of Annual Maxima, Peaks-Over-Threshold and Multifractal Parametric Approaches in Estimating Intensity-Duration-Frequency Curves from Short Rainfall Records." *Journal of Hydrology* 589:125151. https://doi.org/10.1016/j.jhydrol.2020.125151

Fawad, Muhammad, Ting Yan, Lu Chen, Kangdi Huang, and Vijay P Singh. 2019. "Multiparameter Probability Distributions for At-Site Frequency Analysis of Annual Maximum Wind Speed with L-Moments for Parameter Estimation." *Energy* 181:724–737. https://doi.org/10.1016/j.energy.2019.05.153

Fukutome, Sophie, M. A. Liniger, and M. Süveges. 2015. "Automatic Threshold and Run Parameter Selection: A Climatology for Extreme Hourly Precipitation in Switzerland." *Theoretical and Applied Climatology* 120 (3): 403–416. https://doi.org/10.1007/s00704-014-1180-5

Greenwood, J Arthur, J. Maciunas Landwehr, Nicolas C Matalas, and James R Wallis. 1979. "Probability Weighted Moments: Definition and Relation to Parameters of Several Distributions Expressable in Inverse Form." *Water Resources Research* 15 (5): 1049–1054. https://doi.org/10.1029/WR015i005p01049

Haigh, Ivan D, Marta Marcos, Stefan A Talke, Philip L Woodworth, John R Hunter, Ben S Haugh, Arne Arns, Elizabeth Bradshaw, and Phil Thompson. 2023. "GESLA Version 3: A Major Update to the Global Higher-Frequency Sea-Level Dataset." *Geoscience Data Journal* 10 (3): 293–314.https://doi.org/10.1002/gdj3.174

Hamdi, Y, L. Bardet, C.-M. Duluc, and V. Rebour. 2014. "Extreme Storm Surges: A Comparative Study of Frequency Analysis Approaches." *Natural Hazards and Earth System Sciences* 14 (8): 2053–2067. https://doi.org/10.5194/nhess-14-2053-2014

Hanel, Martin, and T. Adri Buishand. 2010. "On the Value of Hourly Precipitation Extremes in Regional Climate Model Simulations." *Journal of Hydrology* 393 (3–4): 265–273. https://doi.org/10.1016/j.jhydrol.2010.08.024

Hawkes, Peter J, Daniel Gonzalez-Marco, Agustin Sánchez-Arcilla, and Panayotis Prinos. 2008. "Best Practice for the Estimation of Extremes: A Review." *Journal of Hydraulic Research* 46 (S2): 324–332. https://doi.org/10.1080/00221686.2008.9521965

Hosking, Jonathan RM 1990. "L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics." *Journal of the Royal Statistical Society: Series B (Methodological)* 52 (1): 105–124. https://doi.org/10.1111/j.2517-6161.1990.tb01775.x

Hosking, Jonathan Richard Morley, James R Wallis, and Eric F Wood. 1985. "Estimation of the Generalized Extreme-Value Distribution by the Method of Probability-Weighted Moments." *Technometrics* 27 (3): 251–261. https://doi.org/10.1080/00401706.1985.10488049

Hosseini, S. R, M. Scaioni, and M. Marani. 2020. "Extreme Atlantic Hurricane Probability of Occurrence Through the Metastatistical Extreme Value Distribution." *Geophysical Research Letters* 47 (1): 2019GL086138. https://doi.org/10.1029/2019GL086138

Huang, Wenrui, Sudong Xu, and Soro Nnaji. 2008. "Evaluation of GEV Model for Frequency Analysis of Annual Maximum Water Levels in the Coast of United States." *Ocean Engineering* 35 (11–12): 1132–1147. https://doi.org/10.1016/j.oceaneng.2008.04.010

Klotzbach, Philip J, Steven G Bowen, Roger Pielke, and Michael Bell. 2018. "Continental US Hurricane Landfall Frequency and Associated Damage: Observations and Future Risks." *Bulletin of the American Meteorological Society* 99 (7): 1359–1376. https://doi.org/10.1175/BAMS-D-17-0184.1

Kulp, Scott A, and Benjamin H Strauss. 2019. "New Elevation Data Triple Estimates of Global Vulnerability to Sea-Level Rise and Coastal Flooding." *Nature communications* 10 (1): 1–12. https://doi.org/10.1038/s41467-019-12808-z

Langousis, Andreas, Antonios Mamalakis, Michelangelo Puliga, and Roberto Deidda. 2016. "Threshold Detection for the Generalized Pareto Distribution: Review of Representative Methods and Application to the NOAA NCDC Daily Rainfall Database." *Water Resources Research* 52 (4): 2659–2681. https://doi.org/10.1002/2015WR018502

Lin, Ning, and Kerry Emanuel. 2016. "Grey Swan Tropical Cyclones." *Nature Climate Change* 6 (1): 106–111. https://doi.org/10.1038/nclimate2777

Makkonen, Lasse 2008. "Problems in the Extreme Value Analysis." *Structural safety* 30 (5): 405–419. https://doi.org/10.1016/j.strusafe.2006.12.001

Mann, Henry B 1945. "Nonparametric Tests Against Trend." *Econometrica: Journal of the Econometric Society* 13 (3): 245–259. https://doi.org/10.2307/1907187

Marani, Marco, and Massimiliano Ignaccolo. 2015. "A Metastatistical Approach to Rainfall Extremes." *Advances in Water Resources* 79:121–126. https://doi.org/10.1016/j.advwatres.2015.03.001

Marani, Marco, and Enrico Zorzetto. 2019. "Doubly Stochastic Distributions of Extreme Events." *arXiv preprint arXiv:1902.09862* 1902 (09862): 1–8. https://doi.org/10.48550/arXiv.1902.09862

Marra, Francesco, Efthymios I Nikolopoulos, Emmanouil N Anagnostou, and Efrat Morin. 2018. "Metastatistical Extreme Value Analysis of Hourly Rainfall from Short Records: Estimation of High Quantiles and Impact of

Measurement Errors." *Advances in Water Resources* 117:27–39. https://doi.org/10.1016/j.advwatres.2018.05.001
Martins, Eduardo S, and Jery R Stedinger. 2000. "Generalized Maximum-Likelihood Generalized Extreme-Value Quantile Estimators for Hydrologic Data." *Water Resources Research* 36 (3): 737–744. https://doi.org/10.1029/1999WR900330
Menéndez, Melisa, and Philip L Woodworth. 2010. "Changes in Extreme High Water Levels Based on a Quasi-Global Tide-Gauge Data Set." *Journal of Geophysical Research Oceans* 115 (C10): 115 (C10. https://doi.org/10.1029/2009JC005997
Miniussi, Arianna, Marco Marani, and Gabriele Villarini. 2020. "Metastatistical Extreme Value Distribution Applied to Floods Across the Continental United States." *Advances in Water Resources* 136:103498. https://doi.org/10.1016/j.advwatres.2019.103498
Mushtaq, Sumra, Arianna Miniussi, Ralf Merz, and Stefano Basso. 2022. "Reliable Estimation of High Floods: A Method to Select the Most Suitable Ordinary Distribution in the Metastatistical Extreme Value Framework." *Advances in Water Resources* 161:104127. https://doi.org/10.1016/j.advwatres.2022.104127
Nasr, Ahmed A, Thomas Wahl, Md Mamunur Rashid, Paula Camus, and Ivan D Haigh. 2021. "Assessing the Dependence Structure Between Oceanographic, Fluvial, and Pluvial Flooding Drivers Along the United States Coastline." *Hydrology and Earth System Sciences* 25 (12): 6203–6222. https://doi.org/10.5194/hess-25-6203-2021
Needham, Hal F, and Barry D Keim. 2012. "A Storm Surge Database for the US Gulf Coast." *International Journal of Climatology* 32 (14): 2108–2123. https://doi.org/10.1002/joc.2425
Needham, Hal F, and Barry D Keim. 2014. "An Empirical Analysis on the Relationship Between Tropical Cyclone Size and Storm Surge Heights Along the US Gulf Coast." *Earth Interactions* 18 (8): 1–15. https://doi.org/10.1175/2013EI000558.1
Neumann, Barbara, Athanasios T Vafeidis, Juliane Zimmermann, Robert J Nicholls, and L. Kumar. 2015. "Future Coastal Population Growth and Exposure to Sea-Level Rise and Coastal Flooding-A Global Assessment." *PLoS ONE* 10 (3): e0118571. https://doi.org/10.1371/journal.pone.0118571
Pugh, David, and Philip Woodworth. 2014. *Sea-Level Science: Understanding Tides, Surges, Tsunamis and Mean Sea-Level Changes*. Cambridge: Cambridge University Press.
Rashid, Md Mamunur, Hamed Moftakhari, and Hamid Moradkhani. 2024. "Stochastic Simulation of Storm Surge Extremes Along the Contiguous United States Coastlines Using the Max-Stable Process." *Communications Earth & Environment* 5 (1): 39. https://doi.org/10.1038/s43247-024-01206-z
Reiss, Rolf-Dieter, and Michael Thomas. 1997. *Statistical Analysis of Extreme Values*. Vol. 2. Basel: Birkhäuser: Springer.
Schellander, Harald, Alexander Lieb, and Tobias Hell. 2019. "Error Structure of Metastatistical and Generalized Extreme Value Distributions for Modeling Extreme Rainfall in Austria." *Earth and Space Science* 6 (9): 1616–1632. https://doi.org/10.1029/2019EA000557
Serafin, Katherine A, Peter Ruggiero, and Hilary F Stockdon. 2017. "The Relative Contribution of Waves, Tides, and Nontidal Residuals to Extreme Total Water Levels on US West Coast Sandy Beaches." *Geophysical Research Letters* 44 (4): 1839–1847. https://doi.org/10.1002/2016GL071020
Solari, Sebastián, Marta Egüen, María José Polo, and Miguel A Losada. 2017. "Peaks Over Threshold (POT): A Methodology for Automatic Threshold Estimation Using Goodness of Fit P-Value." *Water Resources Research* 53 (4): 2833–2849. https://doi.org/10.1002/2016WR019426
Tadesse, Michael Getachew, Thomas Wahl, Md Mamunur Rashid, Sönke Dangendorf, Alejandra Rodríguez-Enríquez, and Stefan Andreas Talke. 2022. "Long-Term Trends in Storm Surge Climate Derived from an Ensemble of Global Surge Reconstructions." *Scientific reports* 12 (1): 1–18. https://doi.org/10.1038/s41598-022-17099-x
Vidrio-Sahagún, Cuauhtémoc Tonatiuh, and Jianxun He. 2022. "Hydrological Frequency Analysis Under Nonstationarity Using the Metastatistical Approach and Its Simplified Version." *Advances in Water Resources* 166:104244. https://doi.org/10.1016/j.advwatres.2022.104244
Villarini, Gabriele, and James A Smith. 2010. "Flood Peak Distributions for the Eastern United States." *Water Resources Research* 46 (6). https://doi.org/10.1029/2009WR008395
Villarini, Gabriele, James A Smith, Francesco Serinaldi, and Alexandros A Ntelekos. 2011. "Analyses of Seasonal and Annual Maximum Daily Discharge Records for Central Europe." *Journal of Hydrology* 399 (3–4): 299–312. https://doi.org/10.1016/j.jhydrol.2011.01.007
Volpi, Elena, Aldo Fiori, Salvatore Grimaldi, Federico Lombardo, and Demetris Koutsoyiannis. 2019. "Save Hydrological Observations! Return Period Estimation without Data Decimation." *Journal of Hydrology* 571:782–792. https://doi.org/10.1016/j.jhydrol.2019.02.017
Wahl, Thomas, Ivan D Haigh, Robert J Nicholls, Arne Arns, Sonke Dangendorf, Jochen Hinkel, and Aimee BA Slangen. 2017. "Understanding Extreme Sea Levels for Broad-Scale Coastal Impact and Adaptation Analysis." *Nature communications* 8 (1): 1–12. https://doi.org/10.1038/ncomms16075
Wahl, Thomas, Shaleen Jain, Jens Bender, Steven D Meyers, and Mark E Luther. 2015. "Increasing Risk of Compound Flooding from Storm Surge and Rainfall for Major US Cities." *Nature Climate Change* 5 (12): 1093–1097. https://doi.org/10.1038/nclimate2736
Wallace, E. E, R. C. MacArthur, S. Chowdhury, and L. K. Sakumoto. 2005. "FEMA Coastal Flood Hazard Analysis and Mapping Guidelines for the Pacific Coast. *Solutions to Coastal Disasters 2005,* edited byLouise Wallendorf, 247–256. Charleston, South Carolina: FEMA.
Wang, Fan, Wei Shao, Haijun Yu, Guangyuan Kan, Xiaoyan He, Dawei Zhang, Minglei Ren, and Gang Wang. 2020. "Re-Evaluation of the Power of the Mann-Kendall Test for Detecting Monotonic Trends in Hydrometeorological Time Series." *Frontiers in Earth Science* 8:14. https://doi.org/10.3389/feart.2020.00014
Woodworth, Philip L, and David L Blackman. 2004. "Evidence for Systematic Changes in Extreme High Waters Since the Mid-1970s." *Journal of Climate* 17 (6): 1190–1197. https://doi.org/10.1175/1520-0442(2004)017<1190:EFSCIE>2.0.CO;2
Zorzetto, E, G. Botter, and M. Marani. 2016. "On the Emergence of Rainfall Extremes from Ordinary Events." *Geophysical Research Letters* 43 (15): 8076–8082. https://doi.org/10.1002/2016GL069445
Zorzetto, Enrico, and Marco Marani. 2020. "Extreme Value Metastatistical Analysis of Remotely Sensed Rainfall in Ungauged Areas: Spatial Downscaling and Error Modelling." *Advances in Water Resources* 135:103483. https://doi.org/10.1016/j.advwatres.2019.103483